JOURNAL OF MULTIVARIATE ANALYSIS **62**, 204–232 (1997)
ARTICLE NO. MV971681

On the Estimation of a Support Curve of Indeterminate Sharpness

Peter Hall,*[,†] Michael Nussbaum,*[,‡] and Steven E. Stern*

*Australian National University, Canberra, Australia;
[†]CSIRO Division of Mathematics and Statistics, Sydney, Australia;
[‡]Weierstrass Institute, Berlin, Germany

We propose nonparametric methods for estimating the support curve of a bivariate density, when the density decreases at a rate which might vary along the curve. Attention is focused on cases where the rate of decrease is relatively fast, this

that employ only a relatively small number of order statistics at the extremities of the point cloud. In this paper we suggest a new type of estimator, based on projecting onto an axis those data values lying within a thin rectangular strip. Adaptive univariate methods are then applied to the problem of estimating an endpoint of the distribution on the axis. The new method is shown to have theoretically optimal performance in a range of settings. Its numerical properties are explored in a simulation study. © 1997 Academic Press

1. INTRODUCTION

Problems of boundary estimation arise in a range of important but quite unrelated contexts. For example, they occur in so-called "scatterpoint image analysis," of the type addressed by Korostelev and Tsybakov (1993a), (1993b) and Mammen and Tsybakov (1995). They are at the heart of problems in "productivity analysis," where the boundary represents the limit to output (Y) that may be expected for a given level of input (X). These data typically arise as a sequence of pairs (X, Y), each pair often representing a different company, for example the railway companies of Europe (Härdle, Hall, & Simar 1995) or U.S. electric utility companies (Christensen & Greene, 1976; Gijbels, Mammen, Park, & Simar 1996). There is a theoretical upper bound to the level of output, y, that

Received May 31, 1995; revised April 6, 1997.

AMS subject classifications: primary 62G07, 62H05; secondary 62H10.

Key words and phrases: convergence rate, curve estimation, endpoint, order statistic, regular variation, support.

0047-259X/97 $25.00

may be achieved for a given input, x, and the bound may be defined by the function $y = g(x)$. All the data pairs (X, Y) lie below the curve, $\mathscr{C}$, represented by this equation. The curve is called the "frontier" to productivity and is to be estimated from a set of values of (X, Y).

Recent work on this class of problems includes that of Härdle, Park, and Tsybakov (1995). There it is assumed that as the support curve $\mathscr{C}$ is approached from below, the density decreases to zero at a constant, known rate. Now, the performance of the curve estimator depends significantly on this rate. If the rate is sufficiently slow then optimal estimation will be based on a relatively small number of bivariate data values at the extremities of the data set. However, if the rate is unknown and fast then optimal estimation can be significantly more difficult and may have to be based on an increasingly large number of bivariate order statistics. (Here and below we adopt the convention that the function u^α, defined for $u > 0$, has a *slow* rate of decrease to zero as $u \downarrow 0$ if $\alpha > 0$ is small, and a *fast* rate of decrease if $\alpha > 0$ is large.)

Our paper addresses precisely the latter context. In contradistinction to previous work on the subject, we (a) assume that the rate of convergence of the density to zero is relatively fast, (b) allow the rate to vary with location along the curve, and (c) do not assume that the rate is known. Thus, we suppose that at a particular point P on $\mathscr{C}$, the density decreases to zero at rate u^α as P is approached from a location distant u below $\mathscr{C}$, where α may be arbitrarily large and depends on P. Our approach to the problem is nonparametric in character, in that we suppose that information about α and the function describing the locus of $\mathscr{C}$ is available only in the form of smoothness conditions, not parametrically.

Even in the one-dimensional case, of estimating the endpoint θ of a distribution, the form of an estimator of θ depends critically on the value of α. We modify a method suggested by Hall (1982) in one dimension, where α is estimated implicitly rather than explicitly. The modification is based on sliding a thin rectangular window through the data. The window is centred on an axis through a point P at which the curve is to be estimated, and those data pairs within the window are projected onto the axis. The estimate at P is then obtained by applying adaptive, univariate methods to the distribution on the axis.

We shall show that this approach produces consistency whenever $\alpha > 1$, and optimal convergence rates in a range of settings when $\alpha > 2$, although not when $1 < \alpha \leqslant 2$. In one dimension, $\alpha = 1$ forms the boundary between cases where it is optimal to use a fixed number of order statistics (corresponding to $\alpha < 1$), and those where optimality demands an increasingly large number of extremes ($\alpha > 1$). Our methods for dealing with the bivariate case are strongly influenced by those for one dimension, and so our interest is also in $\alpha > 1$.

Alternative procedures will produce optimal rates in the latter range, and also in other settings. But in the case where α varies with location, which is the subject of this paper, they are awkward to implement and so are not addressed here. In the case of one dimension, Csörgő and Mason (1989) suggest a two-stage, explicitly defined estimator whose first-order performance is identical to that of Hall's (1982) approach. However, combining the Csörgő–Mason technique with our projection method does not alter convergence rates and has the drawback that it is awkward to apply to bivariate data.

Härdle, Park, and Tsybakov (1995) treat the case of fixed $\alpha \geqslant 0$, but employ estimators based on only a small number of extreme order statistics. Their definition of optimality is different from ours, being based on function classes that provide bounds only to first-order behaviour at the boundary. By way of contrast, our function classes are designed to produce bounds to second-order behaviour. The different convergence rates of estimators that use differing numbers of extreme order statistics do not emerge from Härdle, Park, and Tsybakov's (1995) work.

Section 2 will introduce our methods and describe their main theoretical and numerical properties. Optimal bounds for convergence rates will be presented and derived in Section 3 and shown to coincide in many instances with the rates achieved by the estimators suggested in Section 2. Section 4 will present technical arguments behind the main result in Section 2.

2. ESTIMATION PROCEDURE AND ITS PROPERTIES

We present here our estimator and discuss some of its basic properties. Section 2.1 introduces the basic methodology and describes the actual estimation procedure. Section 2.2 presents our main result concerning asymptotic properties, and Section 2.3 provides discussion and generalisation. Finally, Section 2.4 summarises the results of two simulation studies examining numerical properties of the estimator.

2.1. *Methodology*

There are at least two ways of modelling the distribution of data pairs (X_i, Y_i) below the boundary. First, we may assume that there are just n pairs, and that they are distributed independently according to a bivariate probability density f. Second, we may suppose that the pairs represent points of a Poisson process in the plane, having a bivariate intensity function nf, where f is a fixed nonnegative function. In each case, asymptotic theory involves letting $n \to \infty$ while keeping f fixed. (In the context of Poisson-distributed points, n need not be an integer.) For brevity we shall treat only the former case, of n independently distributed points, in detail. Results in the latter case will be noted briefly in Remark 2.9.

Let $y = g(x)$ denote the locus of the curve above which the bivariate density f is zero. We assume that as $u \uparrow g(x)$, $f(x, y) \to 0$ at an algebraic rate depending on x. Specifically, we suppose that for univariate functions a, b, α, and β, and a bivariate function c,

$$f(x, y) = a(x)\{g(x) - y\}_+^{\alpha(x)} + b(x)\{g(x) - y\}_+^{\beta(x)} + c(x, y)\{g(x) - y\}_+^{\beta(x)} \quad \text{for} \quad x \in \mathscr{I}, \tag{2.1}$$

where $\mathscr{I}$ is a compact interval,

$$a > 0,\ |b| > 0,\ \alpha > 1,\ \beta > \alpha;\ \sup_{x \in \mathscr{I}} |c\{x, g(x) - y\}| \to 0 \quad \text{as} \quad y \downarrow 0;$$

$$\text{the derivatives } a',\ g', \text{ and } \alpha' \text{ exist and are Hölder continuous with exponent } t, \text{ where } 0 \leqslant t \leqslant 1; \text{ and } b,\ \beta \text{ are Hölder continuous.} \tag{2.2}$$

We suppose too that

the marginal density e of X is differentiable, and the derivative is Hölder continuous with exponent t. (2.3)

Next we suggest an estimator of g. Without loss of generality, suppose we wish to calculate $g(0)$ and that 0 is an interior point of $\mathscr{I}$. Given $h > 0$, let (X'_i, Y'_i), for $1 \leqslant i \leqslant N$, denote those data pairs (X_i, Y_i) such that $X_i \in (-h, h)$ are indexed in random order. Write $Y'_{(1)} \leqslant \cdots \leqslant Y'_{(N)}$ for the corresponding order statistics, and following Hall (1982), define

$$\xi_i(\theta) = (Y'_{(N-i+1)} - Y'_{(N-r+1)})/(\theta - Y'_{(N-i+1)}).$$

Our estimator $\hat{g}(0)$ is based on the r largest order statistics, $Y'_{(N-i)}$ for $0 \leqslant i \leqslant r-1$. It is defined to equal the largest solution, θ, of the equation

$$\left[\sum_{i=1}^{r-1} \log\{1 + \xi_i(\theta)\}\right]^{-1} - \left\{\sum_{i=1}^{r-1} \xi_i(\theta)\right\}^{-1} = r^{-1}, \tag{2.4}$$

or to equal $Y_{(N)}$ if no solution exists. One may show that as either $\theta \to \infty$ or $\theta \to Y_{(N)}$, the left-hand side of (2.4) converges to a limit that is strictly less than r^{-1}. Therefore, since the left-hand side is continuous, (2.4) must have an even number of solutions.

While the curve estimator $\hat{g}$ improves substantially on simpler approximations based on large order statistics, it can fluctuate erratically due to the sparsity of information available about g. Arguably the best way of overcoming this problem is to smooth $\hat{g}$. Section 2.4 will investigate the performance of this approach.

Our approach to estimating g is motivated by the frontier estimation problem in productivity analysis, where the Cartesian coordinates of the data pairs (X, Y) have intrinsic meanings. There, it is inappropriate to rotate the coordinate axis, since a linear combination of input and output does not have a well-defined meaning, particularly as a bound to output. As a result, the fact that our main condition (2.1), as well as our estimator of g, is coordinate-dependent, would be expected in productivity analysis.

Nevertheless, if the coordinate system were rotated to a limited extent, so that the curve $\mathscr{C}$ defined by $y = g(x)$ could will properly defined by a Cartesian equation, then an alternative, rotated version of (2.1) would exist, with new functions $\alpha_1, \beta_1, \ldots$ replacing $\alpha, \beta, \ldots$. An estimator of $\mathscr{C}$ could be defined with respect to the rotated system and the "final" estimator derived by averaging over several such rotations.

2.2. *Main Result*

Our first theorem describes large-sample properties of $\hat{g}(0)$. It provides an expansion of the difference $\hat{g}(0) - g(0)$ into bias and error-about-the-mean terms and describes the sizes of the dominant contributions to each. As a prelude to stating the theorem, put $A \equiv 1/\{\alpha(0)+1\}$, $\gamma \equiv \beta - \alpha$,

$$\sigma \equiv \alpha(0)\{\alpha(0)-1\}^{1/2}\{\alpha(0)+1\}^{A-(1/2)}\{e(0)/a(0)\}^{A},$$

$$c_1 \equiv -\tfrac{2}{3}\alpha(0)^2\{\alpha(0)-2\}^{-1}\{\alpha(0)+1\}^{-A}\{a(0)/e(0)\}^{A}\, g'(0)^2,$$

$$c_2 \equiv -\alpha(0)\{\alpha(0)-1\}\{\alpha(0)+1\}^{A\{\gamma(0)+1\}}\beta(0)^{-1}\{\beta(0)+1\}^{-2} \times \gamma(0)^2\, a(0)^{-A\{\beta(0)+2\}}\, b(0)\, e(0)^{A\{\gamma(0)+1\}},$$

$$c_3 \equiv -\tfrac{1}{6}\alpha(0)^4\{\alpha(0)-1\}\{\alpha(0)+1\}^{-(A+1)}\{a(0)/e(0)\}^{A}\, g'(0)^2.$$

Let Q_1 denote a random variable with the standard normal distribution. In the case $1 < \alpha(0) < 2$, define

$$Q_2 \equiv \sum_{i=1}^{\infty} i^{-2A}\left(\sum_{j=1}^{i} Z_j\right)^{-A},$$

where $Z_1, Z_2, \ldots$ are independent exponential random variables with unit mean, independent of Q_1. Recall that N is of size nh, indeed $N/nh \to c$, where $c \equiv 2e(0)$. We may replace N by cnh in the theorem below, without affecting its validity.

THEOREM 2.1. *Assume that the bivariate density f and marginal density e satisfy conditions (2.1)–(2.3), and that $e(0)>0$. Suppose too that for some $0<\varepsilon<\frac{1}{4}$ and all sufficiently large n,*

$$n^{-(1/2)+\varepsilon} \leqslant h \leqslant n^{-\varepsilon}, \qquad n^{\varepsilon} \leqslant r \leqslant n^{1-\varepsilon} h. \tag{2.5}$$

Then if $\alpha(0)>2$ and $nh^{\alpha(0)+2}/r \to 0$,

$$\begin{aligned}\hat{g}(0)-g(0) = {} & (N/r)^A\, h^2 c_1 + (r/N)^{A\{\gamma(0)+1\}}\, c_2 + (r/N)^A\, r^{-1/2}\sigma Q^{(1)} \\ & + O_p(h^{t+1}) + o_p\{(N/r)^A\, h^2 + (r/N)^{A\{\gamma(0)+1\}} + (r/N)^A\, r^{-1/2}\};\end{aligned}$$

if $\alpha(0)=2$,

$$\begin{aligned}\hat{g}(0)-g(0) = {} & (N/r)^A\, h^2 \log r\, c_3 + (r/N)^{A\{\gamma(0)+1\}}\, c_2 + (r/N)^A\, r^{-1/2}\sigma Q^{(1)} \\ & + O_p(h^{t+1}) + o_p\{(N/r)^A\, h^2 \log r + (r/N)^{A\{\gamma(0)+1\}} \\ & + (r/N)^A\, r^{-1/2}\};\end{aligned}$$

and if $1<\alpha(0)<2$ and $nh^{\alpha(0)+2} \to \infty$,

$$\begin{aligned}\hat{g}(0)-g(0) = {} & (r/N)^{A\{\gamma(0)+1\}}\, c_2 + (r/N)^A\, r^{-1/2}\sigma Q^{(1)} + r^{2A-1} N^A h^2 Q^{(2)} c_3 \\ & + O_p(h^{t+1}) + o_p\{(r/N)^{A\{\gamma(0)+1\}} + (r/N)^A\, r^{-1/2} \\ & + r^{2A-1} N^A h^2\},\end{aligned}$$

where $Q^{(1)}$ is asymptotically distributed as Q_1 and, when $\alpha(0)<2$, $(Q^{(1)}, Q^{(2)})$ is asymptotically distributed as (Q_1, Q_2).

2.3. *Discussion of Theory*

The remarks below describe implications and generalisations of Theorem 2.1. If $p(n)$, $q(n)$ are sequences of positive numbers, the notation $p(n) \asymp q(n)$ indicates that $p(n)/q(n)$ is bounded away from zero and infinity as $n \to \infty$.

Remark 2.1. *Simpler statement of the main results.* The results of the theorem may be stated in slightly weaker form as $\hat{g}(0)-g(0) = O_p\{\varepsilon_n(0)\}$ as $n \to \infty$, where $\varepsilon_n = \varepsilon_{n1} + \cdots + \varepsilon_{n4}$ and we define $\varepsilon_{n2}(x) = (r/nh)^{\{\gamma(x)+1\}/\{\alpha(x)+1\}}$, $\varepsilon_{n3}(x) = (r/nh)^{1/\{\alpha(x)+1\}}\, r^{-1/2}$, $\varepsilon_{n4}(x) = h^{t+1}$, and

$$\varepsilon_{n1}(x) = \begin{cases} (nh/r)^{1/\{\alpha(x)+1\}}\, h^2 & \text{if} \quad \alpha(x)>2,\ nh^{\alpha(x)+2}/r \to 0 \\ (nh/r)^{1/\{\alpha(x)+1\}}\, h^2 \log r & \text{if} \quad \alpha(x)=2 \\ r^{\{1-\alpha(x)\}/\{1+\alpha(x)\}} & \\ \quad \times (nh)^{1/\{\alpha(x)+1\}}\, h^2 & \text{if} \quad 1<\alpha(x)<2,\ nh^{\alpha(x)+2}/r \to \infty. \end{cases} \tag{2.6}$$

These results cover the full range $1 < \alpha(x) < \infty$. The side conditions on h are always satisfied if we choose the bandwidth in a way that is optimal from the viewpoint of rates of convergence; see Remarks 2.4–2.7.

Remark 2.2. *Convergences in L^p metrics.* So that we might discuss convergence rates for different values x, let us choose (for each x in a compact interval $\mathscr{I}$) a bandwidth $h(x)$ and an integer $r(x)$ that satisfy $0 < h(x) \leqslant (\log n)^{-1}$, $r(x) \geqslant \log n$, and

$$\begin{aligned} &nh(x)^{\alpha(x)+2}/r(x) \leqslant (\log n)^{-1} && \text{if} \quad \alpha(x) \geqslant 2, \\ &nh(x)^{\alpha(x)+2} > \log n && \text{if} \quad 1 < \alpha(x) \leqslant 2. \end{aligned}$$

(These conditions make redundant the assumptions on h and r in (2.6).) Use this h and r in the definitions of both $\hat{g}(x)$ and $\varepsilon_n(x)$. Assume that the regularity conditions (2.1)–(2.3) hold uniformly in $x \in \mathscr{I}$. Then, minor modifications of the argument employed to prove Theorem 2.1 may be used to show that for each p, $C > 0$,

$$\sup_{x \in \mathscr{I}} \varepsilon_n(x)^{-p} E[\min\{|\hat{g}(x) - g(x)|^p, C\}] = O(1).$$

Therefore, if we modify $\hat{g}$ to $\tilde{g}$ so that it does not take extremely large values (for example, by considering $\tilde{g}(x) = \hat{g}(x)$ if $\hat{g}(x) \leqslant C$ and $\tilde{g}(x) = C$ otherwise, where C is any upper bound to $\sup_{\mathscr{I}} g$), then

$$\int_{\mathscr{I}} E(|\tilde{g} - g|^p) = O\left(\int_{\mathscr{I}} \varepsilon_n^p\right). \tag{2.7}$$

More refined results, for example, with the left-hand side of (2.7) asymptotic to a specific constant multiple of $\int_{\mathscr{I}} \varepsilon_n^p$, may be established using similar arguments. Global performance of $\hat{g}$ will be addressed in the simulation study in Section 2.4.

Remark 2.3. *Sign of bias terms.* Since the constants c_1, c_2, c_3 are all negative then the dominant contributions to the bias of $\hat{g}$ are also negative. In this sense, $\hat{g}$ tends to underestimate g.

Remark 2.4. *Optimal choice of h and r when $\alpha(0) > 2$.* In this range of α there are two deterministic bias terms, of sizes $(N/r)^A h^2$ and $(r/N)^{A\{\gamma(0)+1\}}$, respectively, and one stochastic term describing the error about the mean, of size $(r/N)^A r^{-1/2}$. Recalling that $N \sim cnh$ we see that these three sources of error are of identical size when

$$h \asymp n^{-(\gamma+2)/(2\alpha+5\gamma+4)}, \qquad r \asymp n^{4\gamma/(2\alpha+5\gamma+4)}. \tag{2.8}$$

If $t \geqslant \gamma/(\gamma+2)$ then, with this choice of h and r, the theorem implies that $\hat{g} - g = O_p(\delta_n)$, where $\delta_n \equiv n^{-2(\gamma+1)/(2\alpha+5\gamma+4)}$. It also follows from the theorem that for this choice of h and r, and for t strictly greater than $\gamma/(\gamma+2)$, the limiting distribution of $(\hat{g}-g)/\delta_n$ is normal $\mathrm{N}(\mu, \tau^2)$, where $\mu < 0$ and $\tau > 0$. Observe too that when r and h satisfy (2.8), the conditions (2.5) and $nh^{\alpha+2}/r \to 0$ (imposed in the theorem) are both satisfied.

In the context $\alpha(0) > 2$, at least one special case is of particular interest. For large γ, where the model (2.1) is essentially $f(x, y) \equiv a(x)\{g(x) - y\}_+^{\alpha(x)}$, the optimal sizes of h and r are essentially $n^{-1/5}$ and $n^{4/5} \approx N$, respectively. This bandwidth formula may be recognised as the optimal one for estimation for a twice-differentiable curve. The root mean square convergence rate, of approximately $n^{-2/5}$ when γ is large, is also familiar from that setting. Note particularly that, since $\gamma/(\gamma+2) \leqslant t \leqslant 1$, then $t \to 1$ as $\gamma \to \infty$, and so for large γ we effectively require $t+1=2$ derivatives of g.

For values of t that do not exceed $\gamma/(\gamma+2)$, the optimal convergence rate is achieved not so much by balancing the terms in $(N/r)^A h^2$, $(r/N)^{A\{\gamma(0)+1\}}$, and $(r/N)^A r^{-1/2}$ on the right-hand side of the expansion of $\hat{g}-g$, but by balancing the terms in h^{t+1}, $(r/N)^{A\{\gamma(0)+1\}}$, and $(r/N)^A r^{-1/2}$. Indeed, the theorem implies that when $\alpha(0) > 0$ and we choose

$$
\begin{aligned}
h &= n^{-(\gamma+1)/\{(t+1)(\alpha+2\gamma+1)+\gamma+1\}}, \\
r &= n^{2(t+1)\gamma/\{(t+1)(\alpha+2\gamma+1)+\gamma+1\}}
\end{aligned}
\tag{2.9}
$$

then $\hat{g} - g = O_p(\delta_n)$, where

$$
\delta_n \equiv n^{-(t+1)(\gamma+1)/\{(t+1)(\alpha+2\gamma+1)+\gamma+1\}}. \tag{2.10}
$$

Remark 2.7 will address such results in detail.

Remark 2.5. *Optimal choice of h and r when* $\alpha(0) = 2$. This case is similar to that in the previous remark, with the optimal sizes of h and r differing only by a logarithmic factor from what they were there:

$$
h \asymp \{n^{-(\gamma+2)}(\log n)^{-(\alpha+2\gamma+1)}\}^{1/(2\alpha+5\gamma+4)}, \qquad r \asymp (n^2/\log n)^{2\gamma/(2\alpha+5\gamma+4)}.
$$

If $t > \gamma/(\gamma+2)$ then for these choices of h and r, $\hat{g} - g = O_p(\delta_n)$, where $\delta_n \equiv (n^2/\log n)^{-(\gamma+1)/(2\alpha+5\gamma+4)}$. Indeed, the limiting distribution of $(\hat{g}-g)/\delta_n$ is normal with negative mean and nonzero variance. If $t < \gamma/(\gamma+2)$ then, for h and r chosen according to (2.9), result (2.10) holds.

Remark 2.6. *Optimal choice of h and r when* $1 < \alpha(0) < 2$. The situation here is distinctly different from that when $\alpha(0) \geqslant 2$, in that a new stochastic term with a non-normal asymptotic distribution is introduced into the expansion of $\hat{g}-g$. The optimal sizes of h and r are now

$$h \asymp n^{-(2\alpha+5\gamma+2-\alpha\gamma)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)},$$

$$r \asymp n^{4\gamma(\alpha+1)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)}.$$

(Again, the side condition on h imposed in Theorem 2.1 holds trivially.) If t is sufficiently far from 0 then for such h and r we have $\hat{g}-g=O_p(\delta_n)$, where $\delta_n \equiv n^{-2(\alpha+1)(\gamma+1)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)}$. The asymptotic distribution of $(\hat{g}-g)/\delta_n$ is well defined and representable as a mixture of the distributions of Q_1 and Q_2, together with a location constant.

Remark 2.7. *Optimal convergence rates.* The "optimality" discussed in Remarks 2.4–2.6 is, of course, with respect to choice of tuning parameters for the specific estimator $\hat{g}$ and not necessarily with respect to performance of $\hat{g}$ among all possible approaches. It will turn out, however, that when $\alpha(0)>2$ the convergence rates derived in Remark 2.4 are optimal in the problem of estimating g when the derivative of that function satisfies a Lipschitz condition with exponent $t \leqslant \gamma/(\gamma+2)$. This and related results will be elucidated in the next section.

Indeed, the techniques that we shall employ to derive Theorem 2.1 may be used to obtain the result below, which provides an upper bound to complement the lower bound that will be derived in Section 3. It describes convergence rates of the estimator $\hat{g}$ uniformly over a class of densities more general than those satisfying (2.1)–(2.3). (These stronger conditions are necessary to derive concise expressions for bias and error-about-the-mean terms in Theorem 2.1. However, if only an order-of-magnitude version of that theorem is required, then milder assumptions are adequate.) Let $C>1$ denote a large positive constant, put $\mathscr{J}=[-1/C, 1/C]$, and assume that for univariate functions a, α, and β, and a bivariate function b, the following conditions hold: the density f of (X, Y) satisfies

$$f(x, y)=a(x)\{g(x)-y\}_+^{\alpha(x)}+b(x, y)\{g(x)-y\}_+^{\beta(x)} \qquad \text{for} \quad x \in \mathscr{J},$$

where $C^{-1} \leqslant a \leqslant C$, $|b| \leqslant C$, $2+C^{-1} \leqslant \alpha \leqslant C$, $\alpha+C^{-1} \leqslant \beta \leqslant C$. The derivatives a', g', and α' exist and, denoted by l, satisfy $|l(0)| \leqslant C$ and $|l(u)-l(v)| \leqslant C\,|u-v|^t$ for $u, v \in \mathscr{J}$, where $0 \leqslant t \leqslant 1$; $|\beta(u)-\beta(v)| \leqslant C\,|u-v|^{1/C}$ for $u, v \in \mathscr{J}$. The marginal density e of X is differentiable, $e(0) \geqslant C^{-1}$, $|e'(0)| \leqslant C$, and $|e'(u)-e'(v)| \leqslant C\,|u-v|^{1/C}$ for $u, v \in \mathscr{J}$. Let $\mathscr{F}(t, C)$ denote the class of all such f's.

THEOREM 2.2. *Let h and r be given by* (2.9), *and define δ_n by* (2.10), *in which formulae the functions α and $\gamma=\beta-\alpha$ should be evaluated at the* *origin. Fix* $t \in (0, 1)$. *Then, for all C's which are so large that* $\mathscr{F}(t, C)$ *contains at least one element for which* $\gamma(0)/\{\gamma(0)+2\} \geqslant t$, *we have*

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \sup_{f \in \mathscr{F}(t, C):\, \gamma(0)/\{\gamma(0)+2\} \geqslant t} P\{|\hat{g}(0) - g(0)| \geqslant \lambda \delta_n\} = 0.$$

Remark 2.8. *Alternative estimators of g.* There are several estimators of g alternative to those treated here. In the case where α is known and fixed, estimation may be based on fitting, by maximum likelihood, local, or piecewise polynomials to a and g in the fictitious model $f(x, y) = a(x)\{g(x) - y\}_+^\alpha$. This approach is feasible when the polynomials are linear, but it is not as attractive from a computational viewpoint as the reduction-to-one-dimension method studied in the present paper. The case of second or higher degree polynomials is particularly cumbersome. When α is allowed to vary, a local or piecewise polynomial approximation to that function may be introduced, although this does make the methods very awkward.

The performance of such methods under the more plausible model (2.1) may be described using arguments similar to those developed in Section 4. They attain optimal convergence rates in a wide range of settings, but at the price of significantly increased complexity.

Remark 2.9. *Generalisations to Poisson point processes.* It is straightforward to generalise Theorems 2.1 and 2.2, and also the results in the next section, to the case where the data (X_i, Y_i) originate from a bivariate Poisson processes with intensity λf, where λ is a positive constant. The function f need not be a density, but the only change which that demands is that f need not integrate to 1. The role of n is now played by λ; in particular, the theorems are valid for high-density Poisson processes. In all other respects the conditions required for the theorems remain unchanged. The constants σ, c_1, ..., c_3 defined prior to Theorem 2.1 need to be adjusted, although the c_i's remain negative. With these alterations, Theorems 2.1 and 2.2 hold as before.

Remark 2.10. *Empirical choice of h and r.* Practical implementation of the estimator $\hat{g}$ requires experimentation with different choices of h and r, as shown in Section 2.4. It is often the case that data are too sparse near $\mathscr{C}$ to permit local empirical choice of h, and so global choice is attractive, or perhaps local choice depending on only a small number of tuning parameters (for example, taking $h(x)$ to be linear in x). We suggest employing a simple rule for r, such as that proposed in Section 2.4, and choosing h by experimentation. It may be shown theoretically that a local form of cross-validation generally produces values of h and r of asymptotically correct size, although this result is of questionable practical relevance.

2.4. *Numerical Results*

We present two numerical studies examining the performance of our estimation procedure. The first study addresses the estimator's properties when the boundary is relatively nonlinear. The second examines the estimator's capabilities in distinguishing between a constant boundary with changing exponent function α and a nonconstant boundary with constant exponent α.

In both studies we focus on the case $\alpha(x) > 2$. There, data are distributed extremely sparsely near the boundary, while being plentiful in regions away from the boundary where they have little or no effect on performance. Except in very general terms, sample size is a poor predictor of performance, because it primarily measures data concentration in regions well away from the boundary. Therefore, we give scatter-point plots of typical data sets for each simulation study, so that the reader may determine for himself or herself the difficulty of the estimation problem.

Data pairs (X_i, Y_i) were distributed in the region $g(x) - 2 < y < g(x)$, and the bivariate density $f(x, y)$ was taken equal to a function of x multiplied by $\{g(x) - y\}_+^{\alpha(x)} + \{g(x) - y\}_+^{2\alpha(x)}$, where the function of x was determined so that the distribution of X_i was uniform between 0 and 1. In this situation, $\gamma = \alpha$, and so Remark 2.4 implies that the optimal sizes of the bandwidth and the number of order statistics included in the estimation procedure are $h \asymp n^{-(\alpha+2)/(7\alpha+4)}$ and $r \asymp n^{4\alpha/(7\alpha+4)}$. Using the fact that $N \asymp nh$, it is easily seen that $r \asymp N^{2\alpha/(3\alpha+1)}$, and, therefore, in each of the simulation studies we chose $r(x)$ to be proportional to $\{N(x)\}^{2/3}$.

Simulation Study I. Here we set the boundary curve to be $g(x) = 2 + 4x - 18x^2 + 16x^3$ and the exponent function to be $\alpha(x) = 2 + 3x$ for $x \in [0, 1]$. We chose a sample size of $n = 5000$ points and set $r(x) = 4\{N(x)\}^{2/3}$. Figure 1 shows the results of the new estimation procedure for three different choices of the bandwidth, $h = 0.025, 0.05, 0.1$. The three plots clearly demonstrate the trade-off in variance versus bias as the bandwidth increases. For comparison, each of the plots presents a boundary estimate based solely on the maximum order statistic. Visually, the figures indicate that $\hat{g}$ provides a marked improvement over an estimate based solely on the maximal order statistic. This improvement is borne out in numerical comparisons, too. For example, the mean absolute deviation of the largest order statistic from g, measured within a bandwidth of 0.05 (the context of Fig. 1b), is about 0.58, while that of $\hat{g}$ is approximately 0.30.

One obvious feature of the new estimation procedure is that it produces boundary estimates which are quite "rough" and prone to "spikes." To alleviate this problem it may be useful to consider a variable bandwidth. Alternatively, we might smooth the boundary estimate. Figure 2a presents a LOWESS smooth of the boundary estimate shown in Fig. 1b, as well as boundary estimates using the same bandwidth, h, and number of order

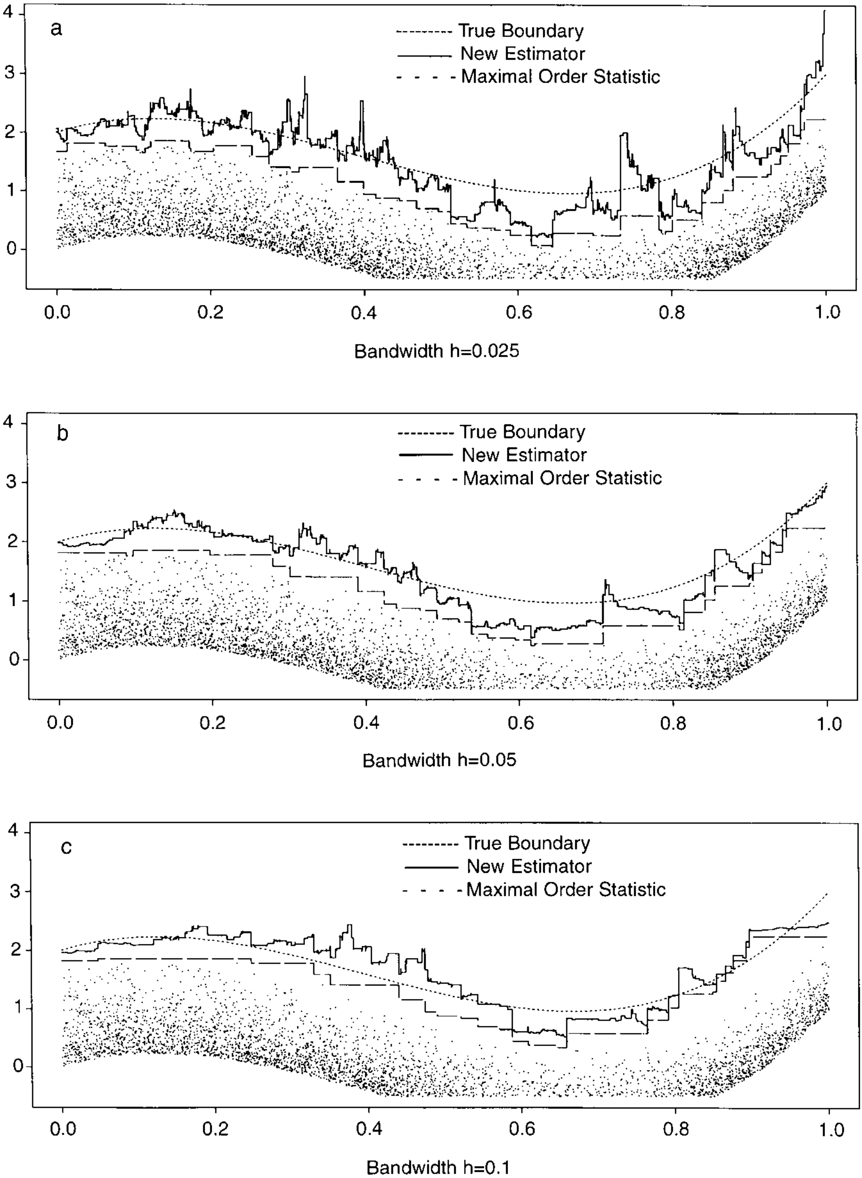
a
True Boundary
New Estimator
Maximal Order Statistic
Bandwidth h=0.025
b
True Boundary
New Estimator
Maximal Order Statistic
Bandwidth h=0.05
c
True Boundary
New Estimator
Maximal Order Statistic
Bandwidth h=0.1

FIGURE 1

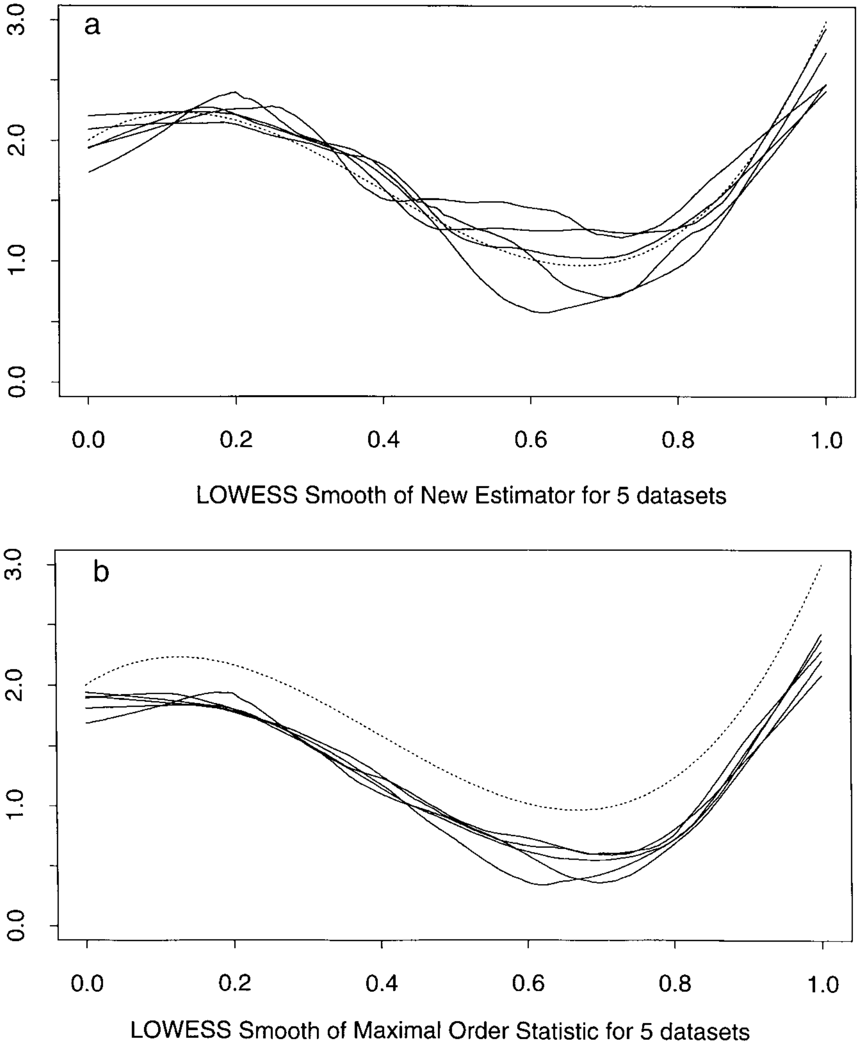

FIGURE 2

statistics, r, for four additional datasets each of size $n = 5000$. Again, for comparison, a LOWESS smooth of the boundary estimates based on the maximal order statistic is presented in Fig. 2b. While the smoothed estimates in Fig. 2b capture the basic shape of the boundary, they are significantly biased. The smoothed version of our new boundary estimate not only captures the shape, but also the location of the boundary, and approximately halves the mean absolute deviation.

Simulation Study II. Here we compare two situations. First, we set the boundary function to be constant, in fact, $g(x) = 2$, and the exponent function to be quadratic, $\alpha(x) = 2 + 24x - 24x^2$ for $x \in [0, 1]$. By way of contrast, in the second situation it is the boundary which is quadratic, $g(x) = 2 - 4x + 4x^2$, while the exponent function is constant at $\alpha(x) = 2$. For samples of size $n = 7500$, each of these two situations produces data which have similar appearances at the upper extremity of the point clouds,

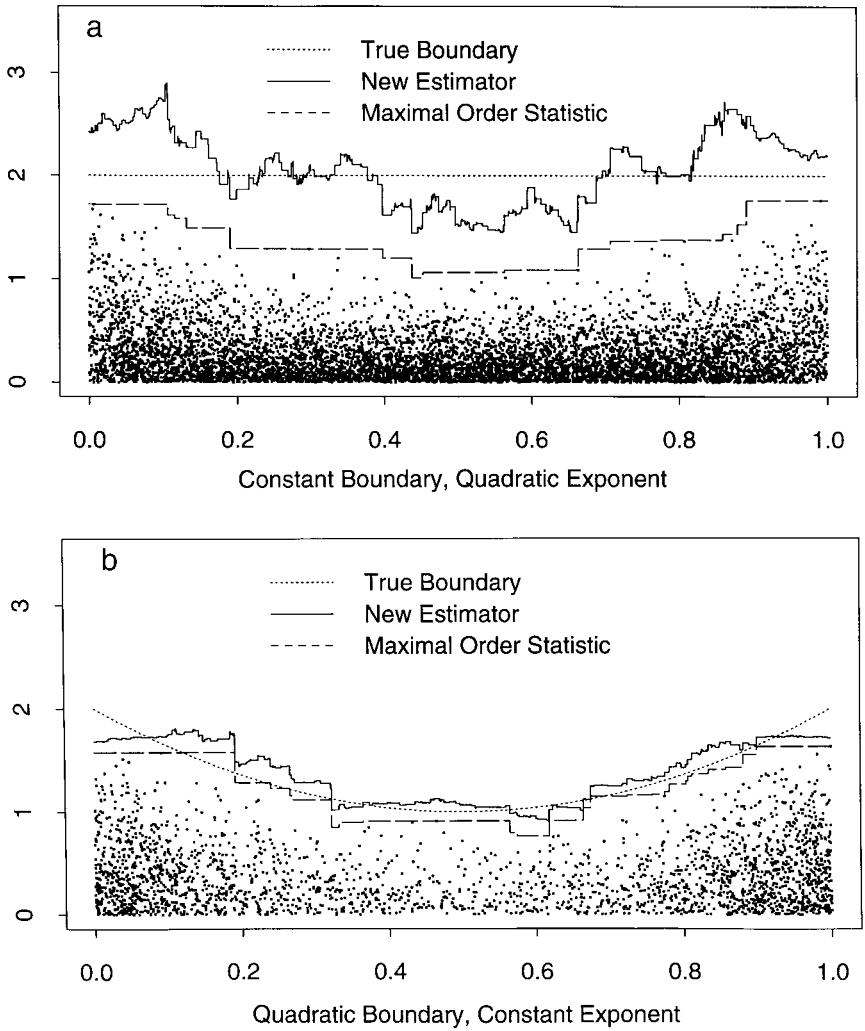

FIGURE 3

despite the difference in boundary curves. This implies that the simple estimator which uses only the largest order statistic within the chosen bandwidth will not be able to easily distinguish between the two situations. However, our estimator, by virtue of its construction using the r largest order statistics, can make the distinction much more readily. Figure 3a presents a plot of the new estimator, as well as the estimator based on the maximal order statistic, in the case of the constant boundary and quadratic

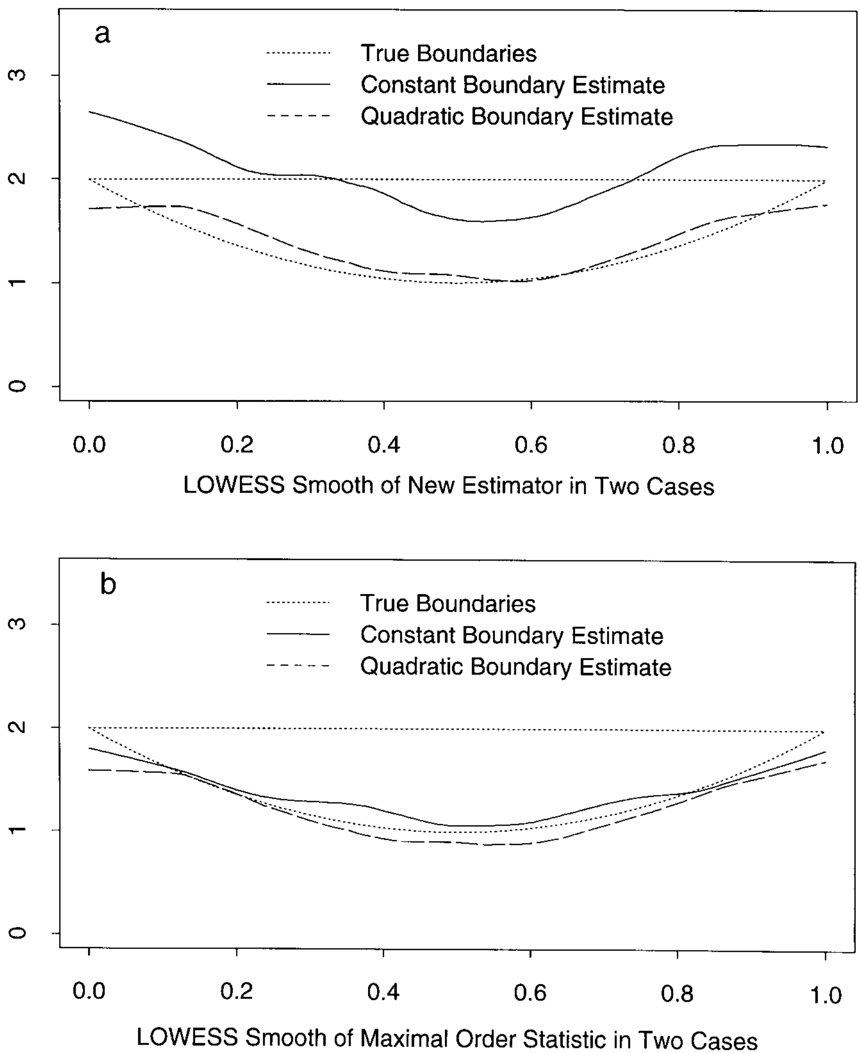

FIGURE 4

exponent function $\alpha(x)$. For this plot the bandwidth was $h = 0.1$, while the number of order statistics used was $r(x) = 8\{N(x)\}^{2/3}$. In contrast, Fig. 3b presents the same estimation procedures in the case of an underlying quadratic boundary with a constant exponent function α. Again, the chosen bandwidth and number of order statistics used are $h = 0.1$ and $r(x) = 8\{N(x)\}^{2/3}$, respectively.

As with the previous simulation study, the new estimation procedure provides somewhat "ragged" curves, although, again, this may mitigated by choice of a more flexible $r(x)$ function or a variable bandwidth. In addition, smoothing may be employed as in the previous example. Figure 4 presents LOWESS smooths of the estimates presented in Fig. 3. Figure 4a shows that the new estimator distinguishes between the two cases to some degree, while Fig. 4b shows that the estimator based solely on the maximal order statistic does not distinguish between the two cases at all.

3. BEST ATTAINABLE CONVERGENCE RATE

In this section we shall assume that the support curve g is of general smoothness $\tau > 0$. More specifically, let $\lfloor \tau \rfloor$ be the largest nonnegative integer $< \tau$ and assume that the derivative $g^{\lfloor \tau \rfloor}$ exists and satisfies $|g^{\lfloor \tau \rfloor}(u) - g^{\lfloor \tau \rfloor}(v)| \leqslant C|u - v|^{\tau - \lfloor \tau \rfloor}$ for $u, v \in \mathscr{I}$. The class of such g's will be denoted by $\Lambda^{\tau}(C)$. For the lower risk bound we shall assume that the functions a, α, and β are known. The assumptions defining the class $\mathscr{F}(t, C)$ in Section 2 will remain in force, with the exception that the lower bound for α will be relaxed to $1 + C^{-1}$, instead of $2 + C^{-1}$. The corresponding class of all f's, when a, α, and β are fixed, will be denoted by $\mathscr{F}'(\tau, C)$. We have to assume that this class is sufficiently rich; there exists $C' < C$ such that $\mathscr{F}'(\tau, C')$ is nonempty.

THEOREM 3.1. *Define δ_n as in (2.10), where $t + 1 = \tau$. Then for all $\tau > 0$,*

$$\lim_{\lambda \to 0} \liminf_{n \to \infty} \inf_{\hat{g}(0)} \sup_{f \in \mathscr{F}'(\tau, C)} P\{|\hat{g}(0) - g(0)| \geqslant \lambda \delta_n\} > 0,$$

where the infimum is taken over all estimators $\hat{g}(0)$ at sample size n.

Introduce notation $A = 1/(\alpha + 1)$, $B = A\gamma$, where $\alpha = \alpha(0)$ and $\gamma = \gamma(0)$, and define a rate exponent ρ by $\delta_n = n^{-\rho}$. In this notation,

$$\rho = \tau/(\tau D^{-1} + 1), \qquad \text{where} \quad D = (A + B)/(2B + 1).$$

To derive Theorem 3.1 we shall follow Donoho and Liu (1991) and consider the value $g(0)$ as a functional on the set of densities f. It is then sufficient to exhibit a sequence of pairs $f_0, f_1 \in \mathscr{F}'(\tau, C)$ such that for the corresponding support curves g_0, g_1 we have

$$H(f_0, f_1) \preccurlyeq n^{-1/2}, \qquad |g_0(0) - g_1(0)| \succcurlyeq n^{-\rho}, \tag{3.1}$$

where $H(\cdot, \cdot)$ is Hellinger distance. In the sequel the notation $n_1 \preccurlyeq n_2$ for two sequences means that $n_1 = O(n_2)$, $n_1 \succcurlyeq n_2$ means that $n_2 = O(n_1)$, and $n_1 \asymp n_2$ means that both $n_1 \preccurlyeq n_2$ and $n_1 \succcurlyeq n_2$. We shall use notation κ (or K) for positive constants, small or large, respectively. The constant C is held fixed at its value in the class $\mathscr{F}'(\tau, C)$.

Consider the problem of endpoint estimation on the real line. Suppose we have random observations Y_i, $i = 1, ..., n$, with density l, where for some $a, \alpha, C > 0$, $\beta > \alpha$, and some θ,

$$l = \bar{l}(\theta - y), \qquad \bar{l}(y) = a y_+^\alpha + b(y)\, y_+^\beta, \qquad |b(y)| \leqslant C. \tag{3.2}$$

Call $\mathscr{F}_0(C)$ the class of densities l in (3.2) when θ varies in R. We shall exhibit a sequence of pairs $l_0, l_1 \in \mathscr{F}_0(C)$ such that for the corresponding endpoints θ_0, θ_1,

$$H(l_0, l_1) \preccurlyeq n^{-1/2}, \qquad |\theta_0 - \theta_1| \succcurlyeq n^{-D}. \tag{3.3}$$

Indeed this will follow from Lemma 3.3 below by putting $\theta \asymp n^{-D}$. For proving (3.3), we will construct, for two given functional values 0 and θ, a pair of densities in $\mathscr{F}_0(C)$ which are at a minimal Hellinger distance. Consider a function

$$l_0(y) = a y^\alpha \qquad \text{for} \quad 0 \leqslant y \leqslant \kappa.$$

Assume κ is small enough so that $l_0(y)$ can be continued to a density outside $[0, \kappa]$. For any $\theta > 0$ define

$$l_1(y, \theta) = \begin{cases} a(y-\theta)_+^\alpha + C(y-\theta)_+^\beta & \text{if} \quad 0 \leqslant y \leqslant y_0(\theta), \\ a y^\alpha & \text{if} \quad y_0(\theta) < y \leqslant \kappa, \end{cases}$$

where the "cutoff point" $y_0(\theta)$ is selected such that

$$\int_0^\kappa l_1(y, \theta)\, dy = \int_0^\kappa l_0(y)\, dy. \tag{3.4}$$

Provided that it is possible, put $l_1(y, \theta) = l_0(y)$ for $y > \kappa$. In view of (3.4), $l_1(\cdot, \theta)$ is also a density. The next two lemmas, whose proofs are omitted, make this precise.

LEMMA 3.1. *For sufficiently small* $\theta > 0$, *a unique solution* $y = \tilde{y}(\theta)$ *of*

$$a(y-\theta)^{\alpha} + C(y-\theta)^{\beta} = ay^{\alpha}, \qquad \theta \leqslant y \leqslant \kappa,$$

exists and satisfies $\tilde{y}(\theta) \sim K_1 \theta^{A/(A+B)}$ *as* $\theta \to 0$, *where*

$$K_1 = \{(A^{-1} - 1)\, aC^{-1}\}^{A/(A+B)}.$$

LEMMA 3.2. *For sufficiently small* $\theta > 0$, *a unique solution* $y = y_0(\theta)$ *of* (3.4) *exists and satisfies* $y_0(\theta) \sim K_2 \theta^{A/(A+B)}$ *as* $\theta \to 0$, *where* $K_2 = \{(B+1)\, A^{-1} aC^{-1}\}^{A/(A+B)}$.

LEMMA 3.3. *As* $\theta \to 0$, $H^2\{l_0, l_1(\cdot, \theta)\} \leqslant K_3 \theta^{1/D}$, *where* $K_3 = aA + K_4 + K_5$, $K_4 = (A^{-1} - 1)^2\, (A^{-1} - 2)^{-1} = a(2K_1)^{1/A - 2}$, $K_5 = K_6^2 Aa(2K_2)^{1/A}$, $K_6 = Ca^{-1}(2K_2)^{B/A}$.

Proof. Define $z(\theta) = \theta^{A/D}$. Consider first the integral from 0 to z. Note that $D = (A+B)/(2B+1) < A+B$. Hence, $z = o(\tilde{y})$, and in this domain we have $l_1(y, \theta) < l_0(y)$. Consequently,

$$\int_0^z \{l_0^{1/2} - l_1^{1/2}(\cdot, \theta)\}^2 \leqslant \int_0^z l_0 = aAz^{1/A} = aA\theta^{1/D}. \tag{3.5}$$

Consider the domain $[z, \tilde{y}]$. Since $A/D = (2AB + A)/(A+B) < 1$, in view of $A < \frac{1}{2}$, we have $y/\theta \to \infty$ uniformly in this domain. Define for $y \in [z, \tilde{y}]$

$$T = l_1(y, \theta)/l_0(y) = [(y-\theta)^{(1/A)-1} + Ca^{-1}(y-\theta)^{\{(B+1)/A\}-1}]/y^{(1/A)-1}.$$

Putting $y = u\theta$, we obtain

$$T = (1-u^{-1})^{(1/A)-1} + Ca^{-1}(1-u^{-1})^{(1/A)-1}\, (u-1)^{B/A}\, \theta^{B/A}. \tag{3.6}$$

By the definition of $\tilde{y}$ we have $T \leqslant 1$ here. Noting that the second term on the right-hand side of (3.6) is positive, we have

$$|1 - T| \leqslant 1 - (1 - u^{-1})^{(1/A)-1}. \tag{3.7}$$

Since $y \geqslant z$ then $1/u = o(1)$ uniformly over $y \geqslant z$. We may, hence, expand the right-hand side in (3.7) and obtain

$$|1 - T| \leqslant u^{-1}(A^{-1} - 1) \sup_{u \geqslant 1}\, (1 - u^{-1})^{(1/A)-2} \leqslant (A^{-1} - 1)\, \theta/y.$$

Here we used again the fact that $(1/A)-2>0$. Evaluating now the integral over this domain, we get

$$\begin{aligned}\int_z^{\tilde{y}} \{l_0^{1/2}-l_1^{1/2}(\cdot,\theta)\}^2 &= \int_z^{\tilde{y}} l_0(1-T^{1/2})^2\\ &\leqslant (A^{-1}-1)\int_z^{\tilde{y}} l_0(y)(\theta/y)^2\,dy\\ &= \theta^2(A^{-1}-1)^2\,a\int_z^{\tilde{y}} y^{(1/A)-3}\,dy\\ &\leqslant (A^{-1}-1)^2\,(A^{-1}-2)^{-1}\,a\theta^2\tilde{y}^{(1/A)-2}.\end{aligned}$$

(Note that $A<\frac{1}{2}$ entails integrability here.) By Lemma 3.1,

$$\int_z^{\tilde{y}} \{l_0^{1/2}-l_1^{1/2}(\cdot,\theta)\}^2 \leqslant K_4\theta^2\theta^{(1-2A)/(A+B)} = K_4\theta^{1/D}, \tag{3.8}$$

where $K_4=(A^{-1}-1)^2\,(A^{-1}-2)^1\,a(2K_1)^{1/A-2}$. The third integral over $[\tilde{y}, y_0]$ will be evaluated as follows. Defining T as in (3.6), we get from the definition of $\tilde{y}$ that $T\geqslant 1$. Then, since the first term on the right-hand side in (3.6) does not exceed 1, $|1-T|\leqslant Ca^{-1}u^{B/A}\theta^{B/A}$. Therefore,

$$|1-T|\leqslant Ca^{-1}(y/\theta)^{B/A}\,\theta^{B/A}\leqslant Ca^{-1}y_0^{B/A}\leqslant Ca^{-1}(2K_2)^{B/A}\,\theta^{B/(A+B)}.$$

Putting $K_6=Ca^{-1}(2K_2)^{B/A}$, we obtain

$$\begin{aligned}\int_{\tilde{y}}^{y_0} \{l_0^{1/2}-l_1^{1/2}(\cdot,\theta)\} &= \int_{\tilde{y}}^{y_0} l_0(1-T^{1/2})^2\\ &\leqslant \theta^{2B/(A+B)}K_6^2\int_{\tilde{y}}^{y_0} l_0\\ &\leqslant K_6^2Aa\theta^{2B/(A+B)}y_0^{1/A}\leqslant K_5\theta^{1/D},\end{aligned}$$

where $K_5=K_6^2Aa(2K_2)^{1/A}$. The lemma follows from this result, (3.5) and (3.8).

In the two-dimensional support curve problem, let f_0 be an element of the class $\mathscr{F}'(\tau, C')$ for a $C'<C$, and let g_0 be the corresponding support curve in the Hölder class $\Lambda^\tau(C')$. Suppose that

$$f_0(x, y)=a(x)\{g_0(x)-y\}_+^{\alpha(x)}+b(x, y)\{g_0(x)-y\}_+^{\beta(x)} \qquad \text{for} \quad x\in\mathscr{I}, \tag{3.9}$$

where $|b(x, y)|\leqslant C'$.

LEMMA 3.4. *The term $b(x, y)$ in (3.9) can be modified such that for some small $\kappa > 0$,*

$$b(x, y) = 0 \qquad \textit{for} \quad 0 \leqslant g_0(x) - y \leqslant \kappa, \ |x| \leqslant \kappa, \tag{3.10}$$

$$|b(x, y)| \leqslant C \qquad \textit{for} \quad x \in \mathscr{I}, \tag{3.11}$$

and the resulting left-hand side in (3.9) is a density in $\mathscr{F}'(\tau, C)$.

Thus, $f_0 \in \mathscr{F}'(\tau, C)$ and the support curve g_0 is in the Hölder class $\Lambda^\tau(C')$. To construct the alternative f_1, let ϕ be an infinitely differentiable function with support in $[-1, 1]$ such that $0 \leqslant \phi(x) \leqslant 1$ and $\phi(0) = 1$. Let $\kappa > 0$ and define a function $\theta(x) = \kappa m^{-\tau} \phi(mx)$, $x \in \mathscr{I}$, where $m > 1$. Define a perturbed support curve g_1 by $g_1(x) = g_0(x) - \theta(x)$, $x \in \mathscr{I}$. This function is in $\Lambda^\tau(C)$ for sufficiently large m if κ is chosen sufficiently small. We shall let m be dependent upon n in the sequel. Specifically, we put

$$m = n^{1/(\tau/D + 1)}. \tag{3.12}$$

LEMMA 3.5. *There is a density $f_1 \in \mathscr{F}'(\tau, C)$ which has support curve g_1 such that $H^2(f_1, f_0) \precsim n^{-1}$.*

Proof. Indicate the dependence of $l_0(y)$ and $l_1(y)$ on θ, a, α, β, C by $l_0(y; a, \alpha, \beta)$ and $l_1(y; \theta, a, \alpha, \beta, C)$. Relations (3.6), (3.11) imply that f_0 can be represented

$$f_0(x, y) = l_0\{g_0(x) - y; a(x), \alpha(x), \beta(x)\} \qquad \text{for} \quad 0 \leqslant g_0(x) - y \leqslant \kappa, \ |x| \leqslant \kappa.$$

Accordingly, define $f_1(x, y) = l_1\{g_0(x) - y; \theta(x), a(x), \alpha(x), \beta(x), C\}$ for $0 \leqslant g_0(x) - y \leqslant \kappa$ and $|x| \leqslant \kappa$, and put $f_1 = f_0$ outside the domain. It follows from (3.9) that for each $x \in \mathscr{I}$, $\int f_1(x, y)\, dy = \int f_0(x, y)\, dy$, so that f_1 is a density which has the same marginal X-density as f_0. By construction of l_1 the density f_1 fulfils

$$f_1(x, y) = a(x)\{g_1(x) - y\}_+^{\alpha(x)} + b(x, y)\{g_1(x) - y\}_+^{\beta(x)} \qquad \text{for} \quad x \in \mathscr{I},$$

where $|b(x, y)| \leqslant C$. We conclude that $f_1 \in \mathscr{F}'(\tau, C)$.

To estimate the Hellinger distance of f_1 from f_0, we argue from Lemma 3.3 and observe that the constants there now depend on x. At this point we need an extension of Lemma 3.3, with uniformity in a, α, β over the range $C^{-1} \leqslant a \leqslant C$, $1 + C^{-1} \leqslant \alpha \leqslant C$, $\alpha + C^{-1} \leqslant \beta \leqslant C$. Such a uniform version can easily be established, on the basis of uniform versions of Lemmas 3.1 and 3.2. With obvious notation, we conclude that $K_3(x)$ is uniformly bounded, while $1/D(x)$ fulfils a Lipschitz condition:

$$|D(x_1)^{-1} - D(x_2)^{-1}| \leqslant K\, |x_1 - x_2|^{1/C}. \tag{3.13}$$

We obtain

$$H^2(f_1, f_0) = \iint \{f_1^{1/2}(x, y) - f_0^{1/2}(x, y)\}^2 \, dy \, dx$$

$$\leqslant \int K(x)\, \theta(x)^{1/D(x)} \, dx \leqslant \kappa K \int \{m^{-\tau}\phi(mx)\}^{1/D(x)} \, dx$$

$$= \kappa K \int_{-\kappa/m}^{\kappa/m} \{m^{-\tau}\phi(mx)\}^{1/D(0)}$$

$$\times \exp[\{D(x)^{-1} - D(0)^{-1}\} \log\{m^{-\tau}\phi(mx)\}] \, dx.$$

Now (3.13) implies that $|D(x)^{-1} - D(0)^{-1}| \leqslant Km^{-1/C}$ so that the term in $\exp(\cdots)$ tends to 0 uniformly in $x \in [-\kappa/m, -\kappa/m]$. Hence,

$$H^2(f_1, f_0) \leqslant K \int_{-\kappa/m}^{\kappa/m} \{m^{-\tau}\phi(mx)\}^{1/D(0)} \, dx$$

$$\asymp m^{-\tau/D(0) - 1} \int \phi(x)^{1/D(0)} \, dx \asymp n^{-1},$$

in view of our selection (3.12) of m. ∎

The respective values of the target functional on f_1 and f_0 are $g_0(0)$ and $g_0(0) - \kappa m^{-\tau}\phi(0)$, so their distance apart is of order $m^{-\tau} \asymp n^{-\tau/(\tau/D + 1)}$. In view of Lemma 3.5, this establishes (3.1).

4. PROOF OF THEOREM 2.1

Observe that for $\eta = \alpha$ or β,

$$\int_{g(0) - u}^{\infty} \{g(x) - y\}_+^{\eta(x)} \, dy = \{\eta(x) + 1\}^{-1} \{g(x) - g(0) + u\}_+^{\eta(x) + 1}. \tag{4.1}$$

If the function ζ is differentiable and ζ' satisfies a Lipschitz condition with exponent t in a neighborhood of the origin, then

$$u^{\zeta(x)} = u^{\zeta(0)}\{1 + x\zeta'(0) \log u + O(x^2 |\log u|^2 + |x|^{t+1} |\log u|)\}, \tag{4.2}$$

uniformly in pairs (x, u) such that $|x \log u|$ is bounded. Put $\zeta = \eta + 1$; let η satisfy the conditions imposed on α in the theorem and let $\lambda = \lambda(h)$ denote a sequence of positive numbers diverging to infinity arbitrarily slowly. Since g' enjoys a Lipschitz condition with exponent t, we have uniformly in $u \in (\lambda h, 1)$ and $|x| \leqslant h$,

$$
\begin{aligned}
&\{g(x)-g(0)+u\}^{\zeta(x)} \\
&\quad =\{u+xg'(0)+O(|x|^{t+1})\}^{\zeta(x)} \\
&\quad =u^{\zeta(x)}[1+u^{-1}x\zeta(x)\, g'(0) \\
&\qquad +\tfrac{1}{2}u^{-2}x^2\zeta(x)\{\zeta(x)-1\}\, g'(0)^2+O(u^{-1}h^{t+1}+u^{-3}h^3)] \\
&\quad =u^{\zeta(0)}[1+u^{-1}x\zeta(0)\, g'(0)+x\zeta'(0)\log u \\
&\qquad +\tfrac{1}{2}u^{-2}x^2\zeta(0)\{\zeta(0)-1\}\, g'(0)^2+O(u^{-1}h^{t+1}+u^{-3}h^3)].
\end{aligned} \tag{4.3}
$$

Therefore, combining (4.1)–(4.3),

$$
\begin{aligned}
&(2h)^{-1}\int_{-h}^{h} dx \int_{g(0)-u}^{\infty} a(x)\{g(x)-y\}_{+}^{\eta(x)}\, dy \\
&\quad =\zeta(0)^{-1}\, a(0)\, u^{\zeta(0)}[1+\tfrac{1}{6}u^{-2}h^2\zeta(0)\{\zeta(0)-1\}\, g'(0)^2 \\
&\qquad +O(u^{-1}h^{t+1}+u^{-3}h^3)].
\end{aligned} \tag{4.4}
$$

Similarly, if η satisfies the conditions imposed on β in the theorem, then

$$
\begin{aligned}
&(2h)^{-1}\int_{-h}^{h} dx \int_{g(0)-u}^{\infty} b(x)\{g(x)-y\}_{+}^{\eta(x)}\, dy \\
&\quad =\zeta(0)^{-1}\, b(0)\, u^{\zeta(0)}[1+O\{(h/u)^{\delta}\}],
\end{aligned} \tag{4.5}
$$

where $\delta > 0$ depends on the exponents of Hölder continuity of b and β. Both (4.4) and (4.5) hold uniformly in $u \in (h^{1-\varepsilon}, 1)$. Furthermore, $P(|X| \leqslant h) = 2h\{e(0)+O(h^{t+1})\}$. Combining this result with (4.4) and (4.5) we deduce that if U has the distribution of $g(0)-Y$ given that $|X| \leqslant h$, then, uniformly in the same range of values of u,

$$
\begin{aligned}
G(u) &\equiv P(U \leqslant u) \\
&= \int_{-h}^{h} dx \int_{g(0)-u}^{\infty} f(x, y)\, dy / P(|X| \leqslant h) \\
&= e(0)^{-1}\, (\{\alpha(0)+1\}^{-1}\, a(0)\, u^{\alpha(0)+1}[1+\tfrac{1}{6}u^{-2}h^2\alpha(0)\{\alpha(0)+1\}\, g'(0)^2] \\
&\quad +\{\beta(0)+1\}^{-1}\, b(0)\, u^{\beta(0)+1}) \\
&\quad +O\{u^{\alpha(0)+1}(u^{-1}h^{t+1}+u^{-3}h^3)+u^{\beta(0)+1-\delta}h^{\delta}\} \\
&= a_1 u^{\alpha(0)+1}\{1+a_2u^{-2}h^2+a_3u^{\gamma(0)}+O(u^{-1}h^{t+1}+u^{-3}h^3+u^{\gamma(0)-\delta}h^{\delta})\},
\end{aligned}
$$

where $\gamma=\beta-\alpha$, $a_1=e(0)^{-1}\{\alpha(0)+1\}^{-1}a(0)$, $a_2=\frac{1}{6}\alpha(0)\{\alpha(0)+1\}\, g'(0)^2$, $a_3=b(0)\{\alpha(0)+1\}/[a(0)\{\beta(0)+1\}]$.

Inverting this expansion we deduce that

$$G^{-1}(v)=b_1 v^{1/\{\alpha(0)+1\}}\{1-b_2 v^{-2/\{\alpha(0)+1\}}h^2-b_3 v^{\gamma(0)/\{\alpha(0)+1\}}$$
$$+O(v^{-1/\{\alpha(0)+1\}}h^{t+1}+v^{-3/\{\alpha(0)+1\}}h^3$$
$$+v^{\gamma(0)/\{\alpha(0)+1\}-\delta}h^{\delta})\}, \tag{4.6}$$

where

$$b_1=[e(0)\{\alpha(0)+1\}/a(0)]^{1/\{\alpha(0)+1\}},$$
$$b_2=\tfrac{1}{6}\alpha(0)[a(0)/e(0)\{\alpha(0)+1\}]^{2/\{\alpha(0)+1\}}\ g'(0)^2,$$
$$b_3=a(0)^{-\{\beta(0)+1\}/\{\alpha(0)+1\}}\ b(0)[\{\alpha(0)+1\}\ e(0)]^{\gamma(0)/\{\alpha(0)+1\}}\ \{\beta(0)+1\}^{-1},$$

uniformly in $v\in(\lambda h^{\alpha(0)+1},\frac{1}{2})$.

Since $g(0)$ is a location parameter, we may assume without loss of generality that $g(0)=0$. In the work below we condition on the value of N, denoting the number of original data pairs (X_i, Y_i) in the interval of width $2h$ centred on the abscissa value $x=0$. Let $U_1, U_2, ..., U_N$ be independent and identically distributed random variables with the distribution of U, and let $U_{(1)}\leqslant U_{(2)}\leqslant \cdots \leqslant U_{(N)}$ denote the corresponding order statistics. In this notation, the sequence $\{\xi_i(\theta), 1\leqslant i\leqslant N\}$ has the same distribution as $\{(U_{(r)}-U_{(i)}/(U_{(i)}-\theta), 1\leqslant i\leqslant N\}$. Without loss of generality, $\xi_i(\theta)=(U_{(r)}-U_{(i)})/(U_{(i)}-\theta)$.

Let $Z_1, ..., Z_N$ denote independent random variables with a common exponential distribution, and define

$$S_i=\sum_{j=1}^{i} Z_j/(N-j+1), \qquad T_i=i^{-1}\sum_{j=1}^{i}(Z_j-1).$$

Noting Rényi's representation for order statistics we see that we may write

$$U_{(i)}=G^{-1}\{1-\exp(-S_i)\}, \qquad 1\leqslant i\leqslant N. \tag{4.7}$$

For any real number w, $S_i=-\log(1-iN^{-1})+(i/N)\{T_i+O_p(i^{1/2}N^{-1})\}$ and

$$\{1-\exp(-S_i)\}^{w}=(i/N)^{w}\{1+wT_i+O_p(i^{-1}+i^{1/2}N^{-1})\}, \tag{4.8}$$

uniformly in $1\leqslant i\leqslant r$.

In the remainder of our proof we treat separately the cases $\alpha(0)>2$, $\alpha(0)=2$, and $1<\alpha(0)<2$. Recall that $A=\{\alpha(0)+1\}^{-1}$.

Case I. $\alpha(0)>2$. Given a positive sequence $\delta(n)\to 0$, let $i_1\geqslant 1$ denote the smallest positive integer such that $(nh/i_1)^A\,h\leqslant\delta(n)$. The assumption

$\varepsilon(n) \equiv nh^{\alpha(0)+2}/r \to 0$, in that part of the theorem dealing with the case $\alpha(0) > 2$, implies that

$$(N/r)^A\, h = O\{\varepsilon(n)^A\}. \tag{4.9}$$

By (4.6)–(4.8) we have, uniformly in $i_1 \leqslant i \leqslant r$,

$$\begin{aligned} b_1^{-1} U_{(i)} = (i/N)^A\, (1 + AT_i - &\{1 + o_p(1)\}\{b_2(N/i)^{2A}\, h^2 + b_3(i/N)^{A\gamma(0)}\} \\ &+ O_p[(N/i)^A\, h^{t+1} + i^{-1} + i^{1/2}N^{-1} \\ &+ \{(N/i)^{2A}\, h^2 + (i/N)^{A\gamma(0)}\}\, i^{-1/2}]). \end{aligned} \tag{4.10}$$

Given a random variable $\tilde{\theta}$ satisfying $N^A\tilde{\theta} \to 0$ in probability, define $\tilde{\theta}_i = (N/i)^A\, \tilde{\theta}$. Put

$$W_1 = r^{-1} \sum_{j=1}^{r} (Z_j - 1) \left\{1 - (1-A)\, r^A \sum_{i=j}^{r} i^{-(A+1)}\right\},$$

$$W_2 = r^{-1} \sum_{j=1}^{r} (Z_j - 1) \left(1 - \sum_{i=j}^{r} i^{-1}\right), \qquad W_3 = (1-A)\, W_1 - W_2,$$

$$d_{11} = (1-2A)^{-1}\, b_1^{-1}, \qquad d_{12} = 2(1-3A)^{-1}\, b_2,$$

$$d_{13} = -\gamma(0)[1 + A(0)\{\gamma(0) - 1\}]^{-1}\, b_3,$$

$$d_{21} = (1-A)^{-1}\, b_1^{-1}, \qquad d_{22} = 2(1-2A)^{-1}\, b_2,$$

$$d_{23} = -\gamma(0)\{A\gamma(0) + 1\}^{-1}\, b_3, \qquad d_{31} = A^2\{(1-A)(1-2A)\}^{-1}\, b_1^{-1},$$

$$d_{32} = 4A^2(1-2A)^{-1}\, (1-3A)^{-1}\, b_2,$$

$$d_{33} = \gamma(0)^2\, A^2(\{1 + A\gamma(0)\}[1 + A\{\gamma(0) - 1\}])^{-1}\, b_3.$$

(Note that, since $\alpha(0) > 2$, $3A < 1$. Also, $d_{3i} = (1-A)\, d_{1i} - d_{2i}$.) In this notation we may prove successively from (4.10) that the following results hold, the first two uniformly in $i_1 \leqslant i \leqslant r$:

$$\begin{aligned} 1 + \xi_i(\tilde{\theta}) &= (U_{(r)} - \tilde{\theta})/(U_{(i)} - \tilde{\theta}) \\ &= (r/i)^A\, [1 + A(T_r - T_i) + b_1^{-1}\{(N/i)^A - (N/r)^A\}\, \tilde{\theta} \\ &\quad + b_2\{(N/i)^{2A} - (N/r)^{2A}\}\, h^2 + b_3\{(i/N)^{A\gamma(0)} - (r/N)^{A\gamma(0)}\} \\ &\quad + O_p[(N/i)^A\, h^{t+1} + i^{-1} + i^{1/2}N^{-1} \\ &\quad + \{(N/i)^{2A}\, h^2 + (i/N)^{A\gamma(0)}\}\, i^{-1/2}] \\ &\quad + o_p\{i^{-1/2} + |\tilde{\theta}_i| + (N/i)^{2A}\, h^2 + (r/N)^{A\gamma(0)}\}\,], \end{aligned} \tag{4.11}$$

$$\begin{aligned}
&\log\{1+\xi_i(\tilde{\theta})\} \\
&\quad = A(\log r - \log i) + A(T_r - T_i) + b_1^{-1}\{(N/i)^A - (N/r)^A\}\,\tilde{\theta} \\
&\qquad + b_2\{(N/i)^{2A} - (N/r)^{2A}\}\,h^2 + b_3\{(i/N)^{A\gamma(0)} - (r/N)^{A\gamma(0)}\} \\
&\qquad + O_p[(N/i)^A\,h^{t+1} + i^{-1} + i^{1/2}N^{-1} \\
&\qquad + \{(N/i)^{2A}\,h^2 + (i/N)^{A\gamma(0)}\}\,i^{-1/2}] \\
&\qquad + o_p\{i^{-1/2} + |\tilde{\theta}_i| + (N/i)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\}\,], \qquad (4.12)
\end{aligned}$$

$$\begin{aligned}
&r^{-1}(1-A)\,A^{-1}\sum_{i=1}^{r-1}\xi_i(\tilde{\theta}) \\
&\quad = 1 + W_1 + d_{11}(N/r)^A\,\tilde{\theta} + d_{12}(N/r)^{2A}\,h^2 + d_{13}(r/N)^{A\gamma(0)} \\
&\qquad + O_p\{(N/r)^A\,h^{t+1} + r^{1/2}N^{-1} + (i_1/r)^{1-A}\} \\
&\qquad + o_p\{r^{-1/2} + (N/r)^A\,|\tilde{\theta}| + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\}, \qquad (4.13)
\end{aligned}$$

$$\begin{aligned}
&r^{-1}A^{-1}\sum_{i=1}^{r-1}\log\{1+\xi_i(\tilde{\theta})\} \\
&\quad = 1 + W_2 + d_{21}(N/r)^A\,\tilde{\theta} + d_{22}(N/r)^{2A}\,h^2 + d_{23}(r/N)^{A\gamma(0)} \\
&\qquad + O_p\{(N/r)^A\,h^{t+1} + r^{1/2}N^{-1} + r^{-1}\log r\} \\
&\qquad + o_p\{r^{-1/2} + (N/r)^A\,|\tilde{\theta}| + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\}. \qquad (4.14)
\end{aligned}$$

(The terms of orders $(i_1/r)^{1-A}$ and $r^{-1}\log r$ on the right-hand sides of (4.13) and (4.15), respectively, derive from extending the sums of the left-hand sides from $i_1 \leqslant i \leqslant r$ (which is their natural range, given the values of i for which (4.11) and (4.12) have been established) to $1 \leqslant i \leqslant r$. For example, in the case of (4.13) observe that $|\xi_i(\tilde{\theta})| = O_p\{(r/i)^A\}$ uniformly in $1 \leqslant i \leqslant i_1$. Hence, the contribution to the left-hand side of (4.13) from such i's is of the same order as the sum of $r^{-1}(r/i)^A$ over those i's. That is, it is of order $(i_1/r)^{1-A}$.)

Therefore,

$$\begin{aligned}
&Ar\left(\left[\sum_{i=1}^{r-1}\log\{1+\xi_i(\theta)\}\right]^{-1} - \left\{\sum_{i=1}^{r-1}\xi_i(\theta)\right\}^{-1} - r^{-1}\right) \\
&\quad = W_3 + d_{31}(N/r)^A\,\tilde{\theta} + d_{32}(N/r)^{2A}\,h^2 + d_{33}(r/N)^{A\gamma(0)} \\
&\qquad + O_p\{(N/r)^A\,h^{t+1} + r^{1/2}N^{-1} + (i_1/r)^{1-A}\} \\
&\qquad + o_p\{r^{-1/2} + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)} + (N/r)^A\,|\tilde{\theta}|\}. \qquad (4.15)
\end{aligned}$$

It follows from (4.15) that if $\tilde{\theta}$ is a solution of Eq. (2.4) then

$$\begin{aligned}-\tilde{\theta} = {} & d_{31}^{-1}(r/N)^{A}\, W_3 + d_{31}^{-1} d_{32}(N/r)^{A}\, h^2 \\ & + d_{31}^{-1} d_{33}(r/N)^{A\{\gamma(0)+1\}} \\ & + O_p[h^{t+1} + (r/N)^{A}\, \{r^{1/2}N^{-1} + (i_1/r)^{1-A}\}] \\ & + o_p\{(r/N)^{A}\, r^{-1/2} + |\tilde{\theta}| + (N/r)^{A}\, h^2 + (r/N)^{A\{\gamma(0)+1\}}\}. \quad (4.16)\end{aligned}$$

Next we show that the term $\tau \equiv O_p[(r/N)^{A}\, \{r^{1/2}N^{-1} + (i_1/r)^{1-A}\}]$, on the right-hand side of (4.16), may be dropped. Since $r/N \to 0$ then $(r/N)^{A}\, r^{1/2}N^{-1} = o\{(r/N)^{A}\, r^{-1/2}\}$, and this term is addressed by the $o_p(\cdots)$ contribution to the right-hand side of (4.16). By definition of i_1, $(i_1/N)^{A} = O\{h\delta(n)^{-1}\}$, and so

$$(i_1/r)^{1-A} = O\{(N/r)^{A}\, h\, \delta(n)^{-1}\}^{(1-A)/A}. \quad (4.17)$$

In view of (4.9) we may choose $\delta(n)$ to converge to zero so slowly that the right-hand side of (4.17) equals $o\{(N/r)^{2A}\, h^2\}$, which is again subsumed into the $o_p(\cdots)$ contribution to the right-hand side of (4.16).

Standard methods may be used to prove that W_3 is asymptotically normally distributed with zero mean and variance $A^2\{r(1-2A)\}^{-1}$. Therefore, defining $\sigma = d_{31}^{-1} A(1-2A)^{-1/2}$, $c_1 = -d_{32}/d_{31}$, and $c_2 = -d_{33}/d_{31}$, we see that from (4.16) (dropping the term corresponding to τ) that

$$\begin{aligned}\tilde{\theta} = {} & (r/N)^{A}\, r^{-1/2}\sigma W_4 + (N/r)^{A}\, h^2 c_1 + (r/N)^{A\{\gamma(0)+1\}} c_2 \\ & + O_p(h^{t+1}) + o_p\{(r/N)^{A}\, r^{-1/2} + (N/r)^{A}\, h^2 + (r/N)^{A\{\gamma(0)+1\}}\}, \quad (4.18)\end{aligned}$$

where W_4 is asymptotically normal N(0, 1). This is equivalent to the claimed expansion in Theorem 2.1. Arguing as in Hall (1982, pp. 566–567) the expansions above may be retraced to show that with probability tending to 1, a solution to (2.4) exists and that the largest solution $\tilde{\theta}$ of (2.4) satisfies $N^{A}\tilde{\theta} \to 0$ in probability. These remarks also apply to the next two cases.

Case II. $\alpha(0) = 2$. Let i_1 be as in Case I, and as before, let $\tilde{\theta}$ denote a random variable equal to $o_p(N^{-A})$. Once again, (4.11) and (4.12) hold uniformly in $i_1 \leqslant i \leqslant r$, and (4.14) is true. In place of (4.13),

$$\begin{aligned} & r^{-1}(1-A)\, A^{-1} \sum_{i=1}^{r-1} \xi_i(\tilde{\theta}) \\ & \quad = 1 + W_1 + d_{11}(N/r)^{A}\, \tilde{\theta} + (1-A)\, A^{-1} b_2 (N/r)^{2A}\, h^2 \log r \\ & \qquad + d_{13}(r/N)^{A\gamma(0)} + O_p\{(N/r)^{A}\, h^{t+1} + r^{1/2}N^{-1} + (i_1/r)^{1-A}\} \\ & \qquad + o_p\{r^{-1/2} + (N/r)^{A}\, |\tilde{\theta}| + (N/r)^{2A}\, h^2 \log n + (r/N)^{A\gamma(0)}\}.\end{aligned}$$

Therefore, (4.15) holds as before but with the term $d_{32}(N/r)^{2A}\, h^2$ replaced by $A^{-1}(1-A)^2\, b_2(N/r)^{2A}\, h^2 \log r$. The analogous change should be made to the right-hand side of (4.16), giving

$$
\begin{aligned}
-\tilde{\theta} = {} & d_{31}^{-1}(r/N)^A\, W_3 + d_{31}^{-1} A^{-1}(1-A)^2\, b_2(N/r)^A\, h^2 \log r \\
& + d_{31}^{-1} d_{33}(r/N)^{A\{\gamma(0)+1\}} \\
& + O_p[h^{t+1} + (r/N)^A\, \{r^{1/2}N^{-1} + (i_1/r)^{1-A}\}] \\
& + o_p\{(r/N)^A\, r^{-1/2} + |\tilde{\theta}| + (N/r)^A\, h^2 \log n + (r/N)^{A\{\gamma(0)+1\}}\}.
\end{aligned}
$$

In view of (4.17) and provided that $\delta(n)$ converges to zero so slowly that

$$\delta(n)(\log n)^{1/2} \to \infty,$$

the term $O_p[(r/N)^A\, \{r^{1/2}N^{-1} + (i_1/r)^{1-A}\}]$ on the right-hand side may be subsumed into the $o_p(\cdots)$ term. Therefore, in place of (4.18),

$$
\begin{aligned}
\tilde{\theta} = {} & (r/N)^A\, r^{-1/2} \sigma W_4 + (N/r)^A\, h^2 \log r\; c_3 + (r/N)^{A\{\gamma(0)+1\}}\; c_2 \\
& + O_p(h^{t+1}) + o_p\{(r/N)^A\, r^{-1/2} + (N/r)^A\, h^2 \log r + (r/N)^{A\{\gamma(0)+1\}}\},
\end{aligned}
$$

where $c_3 \equiv -A^{-1}(1-A)^2\, b_2/d_{31}$. This is equivalent to the claimed expansion in Theorem 2.1.

Case III. $1 < \alpha(0) < 2$. Here it is necessary to develop a refined version of formula (4.11). Our starting point is a more concise form of (4.8) in the special case $w = 1$, which follows via the discussion immediately preceding that result:

$$1 - \exp(-S_i) = (i/N)(1 + T_i)\{1 + O_p(i^{1/2}N^{-1})\},$$

Hence,

$$\{1 - \exp(-S_i)\}^w = (i/N)^w\, \{1 + T_i^{(w)} + O_p(i^{1/2}N^{-1})\}, \tag{4.19}$$

where $T_i^{(w)} \equiv (1 + T_i)^w - 1 = wT_i + O_p(i^{-1})$. Using (4.19) in place of (4.8) we obtain, instead of (4.10), and uniformly in $i_1 \leqslant i \leqslant r$,

$$
\begin{aligned}
b_1^{-1} U_{(i)} = (i/N)^A\, [& 1 + AT_i - \{1 + o_p(1)\}\, b_3(i/N)^{A\gamma(0)} \\
& - \{(1 + T_i^{(A)})(1 + T_i^{(-2A)}) + o_p(1)\}\, b_2(N/i)^{2A}\, h^2 \\
& + O_p\{(N/i)^A\, h^{t+1} + i^{-1} + i^{1/2}N^{-1} + (i/N)^{A\gamma(0)}\, i^{-1/2}\}\,];
\end{aligned}
$$

and in place of (4.11) and (4.13),

$$
\begin{aligned}
1+\xi_i(\tilde{\theta}) &= (U_{(r)}-\tilde{\theta})/(U_{(i)}-\tilde{\theta}) \\
&= (r/i)^A\,[1+A(T_r-T_i)+b_1^{-1}\{(N/i)^A-(N/r)^A\}\,\tilde{\theta} \\
&\quad + b_2(1+T_i^{(A)})(1+T_i^{(-2A)})(N/i)^{2A}\,h^2 \\
&\quad + b_3\{(i/N)^{A\gamma(0)}-(r/N)^{A\gamma(0)}\} + O_p\{(N/r)^{2A}\,h^2 \\
&\quad + (N/i)^A\,h^{t+1}+i^{-1}+i^{1/2}N^{-1}+(i/N)^{A\gamma(0)}\,i^{-1/2}\} \\
&\quad + o_p\{i^{-1/2}+|\tilde{\theta}_i|+(N/i)^{2A}\,h^2+(r/N)^{A\gamma(0)}\}\,],
\end{aligned}
$$

$$
\begin{aligned}
&r^{-1}(1-A)\,A^{-1}\sum_{i=1}^{r-1}\xi_i(\tilde{\theta}) \\
&\quad = 1+W_1+d_{11}(N/r)^A\,\tilde{\theta}+d_{13}(r/N)^{A\gamma(0)} \\
&\qquad + b_2A^{-1}(1-A)\,r^{A-1}N^{2A}h^2\sum_{i=1}^{r-1}(1+T_i^{(A)})(1+T_i^{(-2A)})\,i^{-3A} \\
&\qquad + O_p\{(N/r)^{2A}\,h^2+(N/r)^A\,h^{t+1}+r^{A-1}+r^{1/2}N^{-1}\} \\
&\qquad + o_p\{r^{-1/2}+(N/r)^A\,|\tilde{\theta}|+r^{A-1}N^{2A}h^2+(r/N)^{A\gamma(0)}\}.
\end{aligned}
$$

In view of the assumption $nh^{\alpha(0)+2}\to\infty$, made in that part of the theorem addressing the case $1<\alpha(0)<2$, the term $O_p(r^{A-1})$ is of smaller order than $r^{A-1}N^{2A}h^2$, and so may be incorporated into the remainder $o_p(r^{A-1}N^{2A}h^2)$. Similarly, the $O_p(r^{1/2}N^{-1})$ term is subsumed by the remainder $o_p(r^{-1/2})$. Results (4.12) and (4.14) hold as before. Therefore, instead of (4.18),

$$
\begin{aligned}
\tilde{\theta} &= (r/N)^A\,r^{-1/2}\sigma W_4+r^{2A-1}N^Ah^2W_5+(r/N)^{A\{\gamma(0)+1\}}\,c_2 \\
&\quad + O_p(h^{t+1})+o_p\{(r/N)^A\,r^{-1/2}+r^{2A-1}N^Ah^2 \\
&\quad + (r/N)^{A\{\gamma(0)+1\}}\},
\end{aligned} \tag{4.20}
$$

where

$$
W_5\equiv c_3\sum_{i=1}^{\infty}(1+T_i^{(A)})(1+T_i^{(-2A)})\,i^{-3A},
$$

and c_3 is defined as in the previous case. Result (4.20) is equivalent to the claimed expansion in Theorem 2.1.

ACKNOWLEDGMENTS

The authors are grateful to a referee for helpful comments.

REFERENCES

Christensen, L., and Greene, R. (1976). Economics of scale in U. S. electric power generation. *J. Polit. Econ.* **84** 653–667.

Csörgő, S., and Mason, D. (1989). Simple estimators of the endpoint of a distribution, In *Extreme Value Theory*. Proceedings, Oberwolfach, 1987 (J. Hüsler and R.-D. Reiss, Eds.), Lecture Notes in Statistics, Vol. 51. Springer-Verlag, New York.

Csörgő, S., Deheuvels, P., and Mason, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13** 1050–1077.

Donoho, D. L., and Liu, R. C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633–667.

Gijbels, I., Mammen, E., Park, B. U., and Simar, L. (1996). On estimation of monotone and concave frontier functions. [manuscript]

Härdle, W., Park, B. U., and Tsybakov, A. B. (1995). Estimation of nonsharp support boundaries. *J. Multivariate Anal.* **55** 205–218.

Härdle, W., Hall, P., and Simar, L. (1995). Iterated bootstrap with applications to frontier models. *J. Productivity Anal.* **6** 63–76.

Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

Janssen, A., and Marohn, F. (1994). On statistical information of extreme order statistics, local extreme value alternatives, and Poisson point processes. *J. Multivar. Anal.* **48** 1–41.

Korostelev, A. P., and Tsykabov, A. B. (1992). Efficient support of a probability density and estimation of support functionals. Institut de Statistique, Louvain-la-Neuve. [Discussion paper 9229]

Korostelev, A. P., Simar, L., and Tsykabov, A. B. (1992). Efficient estimation of monotone boundaries. Institut de Statistique, Louvain-la-Neuve. [Discussion paper 9206]

Korostelev, A. P., and Tsykabov, A. B. (1993a). Estimation of support of a probability density and of its functionals. *Problems of Information Transmission* **29** 3–18.

Korostelev, A. P., and Tsykabov, A. B. (1993b). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics, Vol. 82. Springer-Verlag, New York/Berlin.

Mammen, E., and Tsykabov, A. B. (1995). Asymptotical minimax results in image analysis for sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

Marohn, F. (1991). Global sufficiency of extreme order statistics in location models of Weilbull type. *Probab. Theory Related Fields* **88** 261–268.